

# Mercury

## Distributed Metadata Management, Data Discovery, and Access System

An internet tool to access and distribute data

Tammy Walker Beaty

NASA ORNL DAAC  
National Aeronautics and Space Administration  
Oak Ridge National Laboratory  
Distributed Active Archive Center  
Oak Ridge, Tennessee, USA

# About the ORNL DAAC

<http://daac.ornl.gov/>

## ➤ NASA's ORNL DAAC

Distributed Active Archive Center for biogeochemical and ecological data and models useful in environmental research

986 published data sets and counting

4 areas within NASA's Terrestrial Ecology Program

- Field Campaigns, Validation, Regional and Global, Model archive

Web interfaces to ORNL DAAC and NASA's 12 Earth Observing System data centers

Tools for ingesting data and for reporting user activity

# Multiple Distribution Mechanisms

- FTP
- Metadata Search Catalog – Mercury (FGDC compliant)
- NASA Earth Observation System-wide clearinghouse (ECHO/Reverb)
- Spatial Data Access Tool (WCS, WMS)
- WebGIS (WMS)
- THREDDS Data Server
- Tools for subsetting and visualization of remote sensing products (MODIS)

Data are available at no cost

# Diverse Goals

- ORNL DAAC has a diverse customer base
- Education and Outreach - data contributors, research scientists, educators, legislators, students
- Assembling data bases for major resources
- Developing technologies are as important to provide as the source data
- Security issues related to policy and privacy as well as implementation of governing standards must be addressed.

# Data Management Challenges

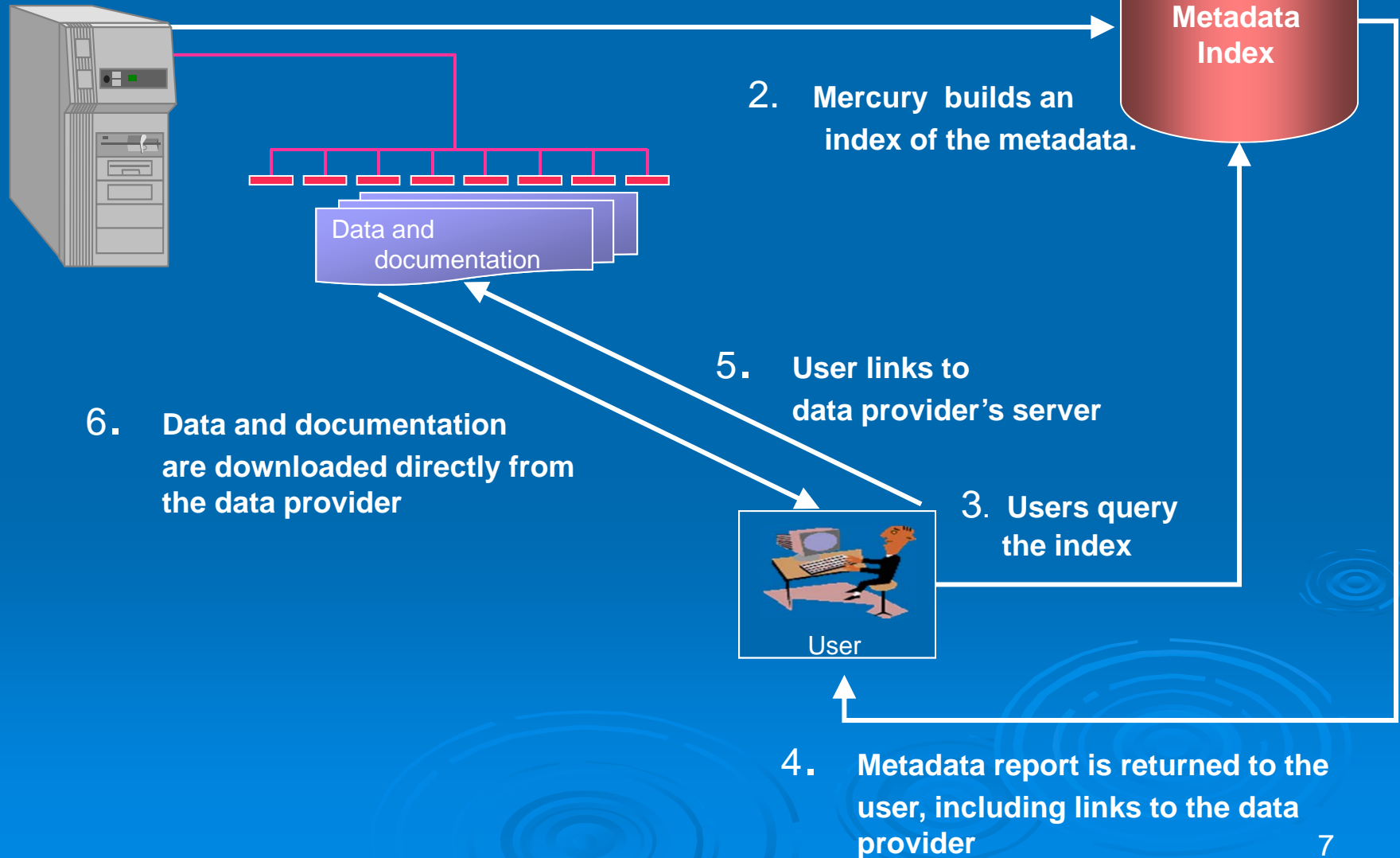
- World full of diverse data, scientists, GIS and DBMS software, custom models, and remote sensing resources.
- Success in science based on publications. Data published in papers, methods generalized, source information cited but usually only results published, not the raw data files.
- You need sophisticated search skills to find exactly what you need through Web search engines like Google.
- Certain data need to be archived with access restrictions in place.
- Similar information in multiple places produced with different tools at different scales.

# What is Mercury ?

- Distributed metadata management, data discovery, and data access system
- Metadata Search and Retrieval tool
- Standards based – XML, Z39.50, FGDC, Dublin-Core, Darwin-Core, EML, and ISO 19115 (development pending)
- Based on open source and Service Oriented Architecture

# Mercury: *Metadata Search and Data Access System*

1. The data provider uses the ORNL Metadata Editor to create a metadata file containing links to the data, documentation, and ancillary files.



# Two Mercury Models

## ➤ The Virtual Internet Database

- Organizes a new collection of data from informal systems spread across the Internet

## ➤ The Virtual Aggregate Database

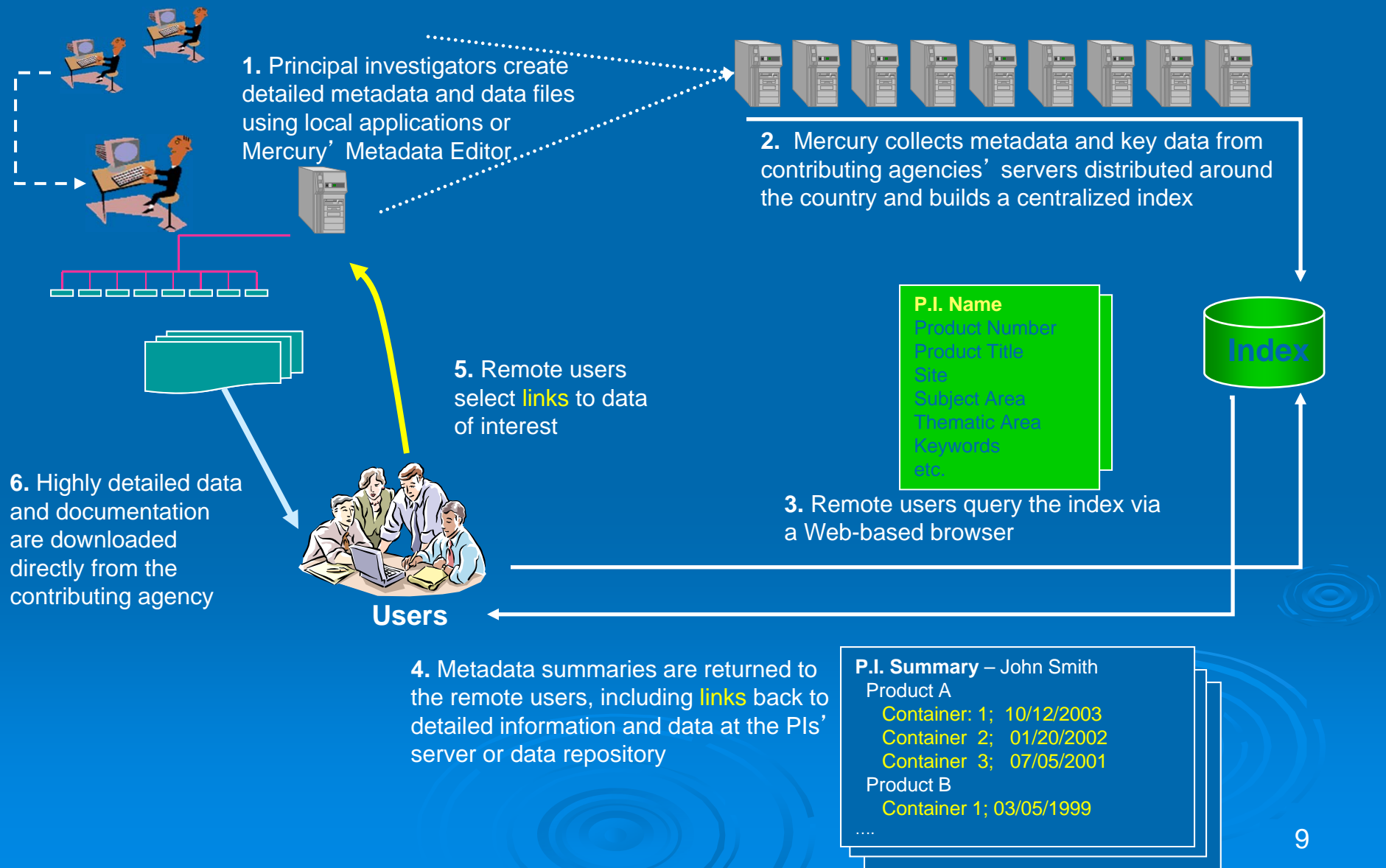
- Harvests information from existing formal separate DBMS

## ➤ And Many Hybrids...

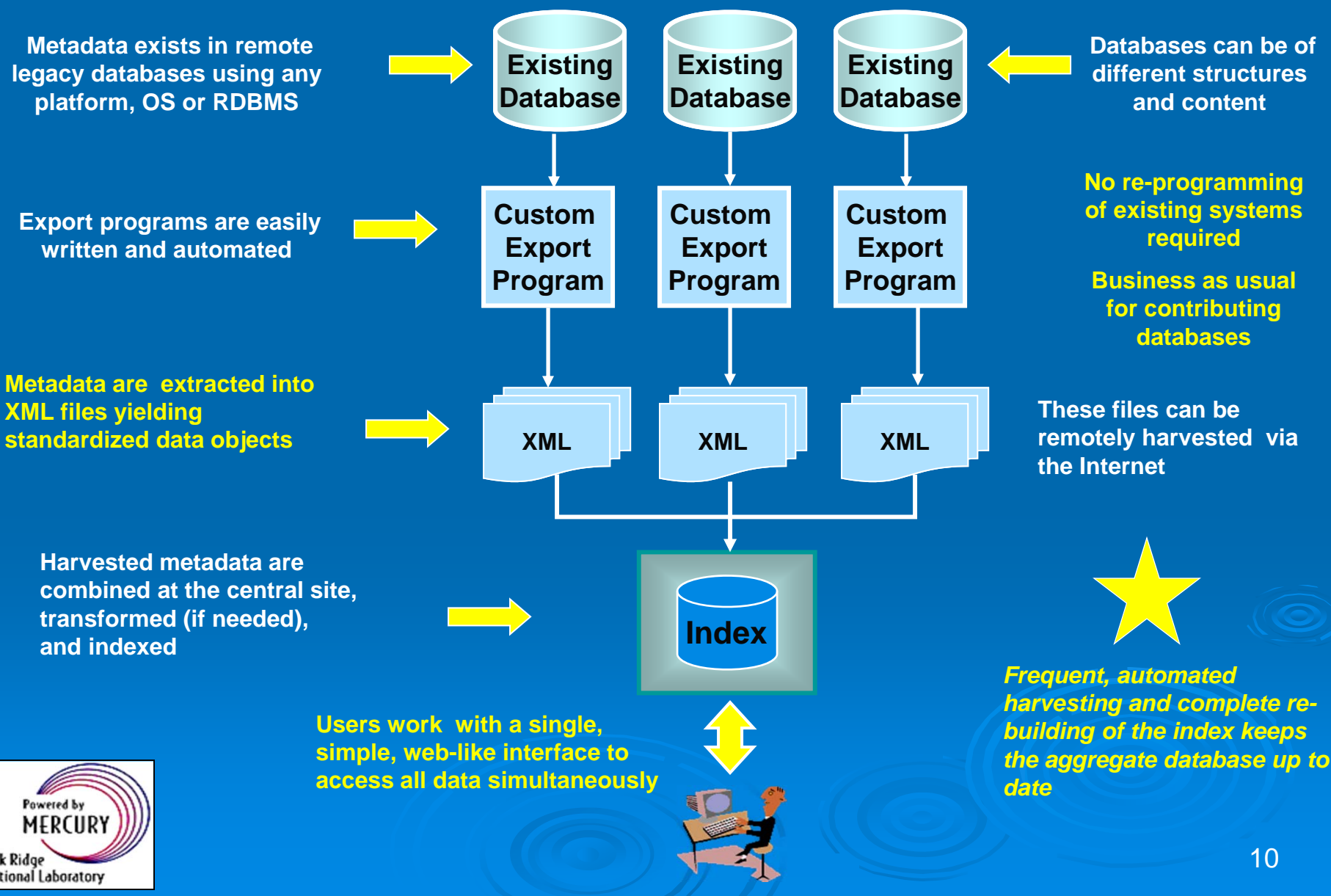


# Mercury : Virtual Internet Database

## Distributed Data Discovery and Access System



# A Virtual Aggregate Database via Mercury



# What Can Mercury Do?

- Provide a **single portal** to information contained in separate data management systems
- Provide **fielded, spatial, and temporal search** capabilities
- Allow individuals and database managers to **distribute their data** while maintaining complete control and ownership
- **Leverage investment** in existing information systems and research

# Custom Interfaces

## LP DAAC (Land Processes, EROS, USGS)

[USGS Home](#)  
[Contact USGS](#)  
[Search USGS](#)

LP DAAC  
 LAND PROCESSES DISTRIBUTED ACTIVE ARCHIVE CENTER

## LP DAAC Metadata Clearinghouse

---

**Simple Search**

Search For:

Results/ page  
**10** ▼

Hint: boolean operators, wildcards and phrases are allowed. ex: precipitation or (rain) and "moisture content"

---

**Advanced Search**

**Fielded Search**

FullText ▼

[Help](#) | [Clear](#)

**Search by Date Range**

during ▼  thru   
mm/dd/yyyy mm/dd/yyyy

[Help](#) | [Clear](#)

---

**Geographic Search**

POWERED BY Google

\*Click on [icon] to select an area

Last Area in:  
USA | WORLD ▼

Select from list ▼

Search Area:

[surface](#) | [undersea](#)

North

West East

South

[Help](#) | [Clear](#)

---

Query being built:

Not Editable

Results/ page  
**10** ▼

---

LAND PROCESSES DISTRIBUTED ACTIVE ARCHIVE CENTER

47914 232nd Street Sioux Falls, SD 57198-0001 Voice: 605-594-8116 Toll Free: 866-573-3222 Fax: 605-594-8962 lpdaac@usgs.gov

[Accessibility](#)
[FOIA](#)
[Privacy](#)
[Policies and Notices](#)

[U.S. Department of the Interior](#) | [U.S. Geological Survey](#)  
 URL: <https://lpdaac.usgs.gov/lpdaac>  
 Page Contact Information: [lpdaac@usgs.gov](mailto:lpdaac@usgs.gov)  
 Page Last Modified: October 07, 2010

# Mercury Demonstration

- Metadata Editor – existing editor created in late 1990s

From: /data/bluangel\_harvest/MERCURY/PD/basic\_example.xml  
(READONLY - cannot be overwritten) (\*=mandatory field)

- Searchable Metadata: (-)
  - o Data Originator: (+) (2/5 items completed)
  - o Data Set Information: (+)
    - Data Set Title: (m s g) Landset 7, Level 1G, ETM+ data
    - Project: (One project per line)(m s g)
      - EOS Land Validation: ETM+ data for the Skukusa EOS Land Validation Core Site: Path 168, Row 77
    - Site Information: (+) (All items completed)
    - Time Period: (+) (2/3 items completed)
    - Parameter Description: (+) (All items completed)
    - Spatial Reference Method: (show pick list) (m s g)
  - Keywords: (One keyword per line)(m s g)
    - Validation
    - EOS Land Validation Core Site
    - EOS Land Validation Core Sites
    - SAPART 2000
  - Abstract: (m s g)
    - ETM+ data acquired on 16 October 1999.
    - ECS Granule ID: E15C:L70RWR9.002:2000316960
    - Ordered through ECS on 12/09/1999
    - Level: 1G
    - Format: GeoTiff
    - Projection: UTM
    - Zone: 36
    - Calibration: CPF
    - Band: all
    - Orientation: NUP
    - Cell size (non/UNIT/Therm): 15/30/60
- o Data Access Information: (+) (10/11 items completed)
- o Alternate Data Access: (+) (0/11 items completed)
- o Online html, xml, or text file to be included in free-text search: (+) (All items completed)
- o Online resource NOT to be included in free-text search: (+) (0/2 items completed)
- Data Set Documentation: (+) (0/7 items completed)

- New Internal and External metadata editor – in development
  - covered by Viv Hutchinson

# Mercury Demonstration

➤ <http://mercury.ornl.gov/ornldaac/index.jsp?tab=advanced>

- 8 data sources
  - ORNL DAAC Archives (986)
  - FLUXNET (548)
  - LBA-ECO (614)
  - RGD (452)
  - Land Val (68)
  - LP DAAC MODIS & ASTER (102)
  - Long Term Ecological Research Network , LTER (7048)
  - Organization of Biological Field Stations, OBFS - 124

# Questions?

# Additional Slides



# Data Processing and Support Challenges in Data Archival

## ❑ Acquisition

- identify how best to serve the audience / scientific community
- establish how to obtain data
- acquire data (local archive) or link to data (mercury, hard link)

## ❑ Ingest – Archival of Data Sets

- perform QA checks
- compile project-provided metadata
- generate additional metadata

## ❑ Enhance (as requested)

- add supplementary variables
- standardize to common units
- aggregate files

## ❑ Documentation

- collect available documentation and write “guide” documents for each data set, as needed

## ❑ Distribution

- advertise data (web, newsletters, customer announcements)
- provide data access to users worldwide

## ❑ Post-Project Data Support

- provide long-term secure archiving of the data
- serve as a buffer between end users and PIs
- provide usage statistics for Sponsors
- survey users about major products and/or projects

## ❑ Stewardship

- security, disaster recovery
- Ensure life of the data (reformat when proprietary software or operating systems reconfigure data formats)

# ORNL DAAC: Data Collections

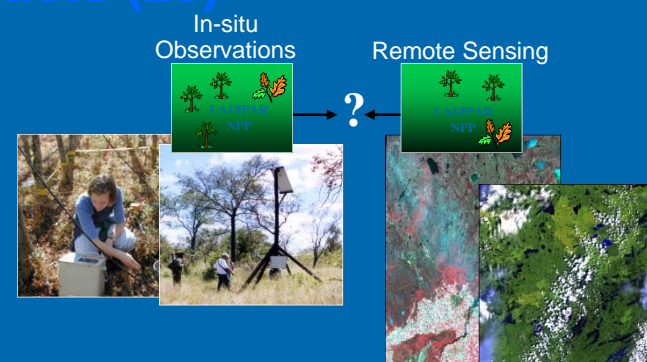
(Number of Data Sets = 986)

## 1. Field Campaigns (755)

- 6-9 year intensive study of a region:
  - Amazon (LBA)
  - Northern Canada (BOREAS)
  - Southern Africa (SAFARI 2000)

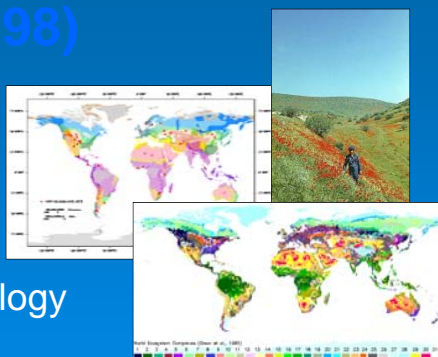


## 2. Validation of Remote Sensing Products (23)



## 3. Regional and Global Studies (198)

- Climate
- Soils
- Vegetation
- Hydroclimatology



## 4. Model Code (10)

- Benchmark Models
  - IBIS, BIOME-BGC, LSM
- Manuscript Models
  - PNeT, Century, Biome-BGC

